# A Novel Molecular Subtyping Based on Proteomics Analysis for Prognosis Predicting in Periprosthetic Joint Infection

Chenghan Chu[1], Baiqi Pan[1], Xiaoyu Wu[1], Yang Xing[1], Lingli Long[1], Puyi Sheng[1]
1The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China
chuchh@mail2.sysu.edu.cn

INTRODUCTION: Periprosthetic joint infection (PJI) is a devastating complication after arthroplasty, with high re-infection rates. Establishing a classification to identify patients at risk for recurrent PJI could assist patient stratification and early precise treatments. Current classifications, based mainly on clinical characteristics, fail to comprehensively assess PJI severity, limiting their predictive power. To better understand PJI biology and accurately stratify patients, we aim to establish a proteomic molecular subtype to identify those at different re-infection risks. We hypothesize that this classification will effectively distinguish high-risk PJI patients and provide insights into the biological mechanisms of PJI.

METHODS: We applied proteomic analysis based on data-independent acquisition-mass method on synovium paraffin tissue of 150 PJIs (61 male, 89 female) from multi-centers between 2015 and 2024. The samples were divided into discovery cohort (100 samples) and validation cohort (50 samples). Besides, we also applied proteomic analysis on plasma of 100 PJIs in discovery cohort. ConsensusClusterPlus clustering was conducted on the tissue of discovery cohort to form molecular subtypes for accessing of re-infection rate and functional analysis. Furthermore, tissue protein signatures were identified using machine leaning algorithms. A random forest model based on these signatures was applied to predict PJI classification in the validation cohort. Besides, the three plasma proteins with the highest AUC values in each subtype are regarded as plasma signatures. Proteomic data were analyzed and visualized using R.

RESULTS SECTION: In the discovery cohort, unsupervised clustering identified three distinct subtypes: S1, S2, and S3 (50:30:20)**(Fig.1A)**. Baseline demographics showed no significant differences among the subtypes. KEGG analysis of differentially expressed proteins (DEPs) revealed that the S1 subtype was enriched in the cytoskeleton pathway, the S2 subtype in immune activation pathways, and the S3 subtype in cholesterol metabolism pathways **(Fig.1B)**. Additionally, the ESTIMATE analysis demonstrated that the S3 subtype had a lower immune score, suggesting an immunosuppressive microenvironment **(Fig.1C)**. Further analysis revealed a strong correlation between molecular subtype and reinfection rates (P=0.034) **(Fig.1D)**. Besides, we recorded the frequency of each operation type selected for physician assessment. Notably, in S1 and S2 subtype, there is no statistically significant difference in the recurrence rates between two-stage revision and one-stage revision **(Fig. 1E-F)**, while in S3 subtype, the recurrence rate for two-stage revision is significantly lower than that for one-stage revision (P=0.045) **(Fig. 1G)**. Furthermore, using machine learning algorithms, we identified a subtype signature comprising CTSG, AZU1, SOAT1, GPNMB, CYFIP2, and SPON1, and developed a random forest prediction model based on the expression levels of these signatures **(Fig.1H)**. We applied the model to validation cohort to derive the subtype assignments, which replicated the associations with re-infection rate (P=0.036) **(Fig.1I)**. Finally, we also conducted plasma proteomics analyses on the three subtypes of patients in the discovery cohort. And three specific plasma signatures of S1 (MMP3, APOBR, PRB1), S2 (RPS24, TPST1, PAK6) and S3 (GPC1,FGL2, OLFML2B) subtypes were screened out (**Fig. 1J-L**).

DISCUSSION: We can identify patients with high re-infection rate based on proteomics molecular subtyping. Functional analysis shows that the cholesterol metabolic-rich subtype demonstrated immune suppression, thereby probably resulting in a higher re-infection rate for individuals. Meanwhile, prognostic analysis indicated that the S3 subtype might be more suitable for two-stage revision due to its lower recurrence rate. Finally, we can classify PJI by detecting synovium or plasma proteins. Our results provide crucial insights for accurate prognosis prediction and preciese treatment of PJI, while help comprehend the molecular mechanisms of PJI.

SIGNIFICANCE/CLINICAL RELEVANCE: (1-2 sentences): Proteomic molecular subtypes can be used to identify patients with high risk of re-infections and assist early precise treatments.

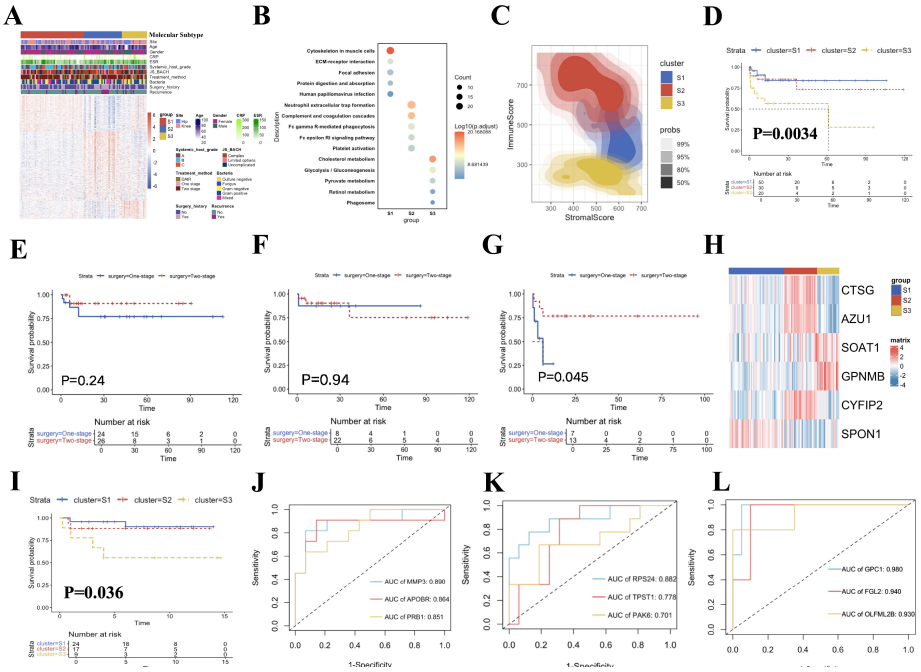


Figure 1: **A.** Three clusters were identified using unsupervised clustering analysis based on proteomics. **B.** KEGG analysis based on DEPs. **C.** Two-dimensional density contour plot based on immune score (y axis) and stromal scores (x axis) for three subtypes. **D.** The Kaplan- Meier curves of three molecular subtypes in discovery cohort. **E.** The recurrence rate of one-stage revision and two-stage revision in patients with S1 subtype. **F.** The recurrence rate of one-stage revision and two-stage revision in patients with S2 subtype. **G.** The recurrence rate of one-stage revision and two-stage revision in patients with S3 subtype. **H.** The expression of six signature proteins for the constitutive prediction model. **I.**The Kaplan-Meier curves of three predicted molecular subtypes in validation cohort. **J.**The sigature proteins of S1subtype. **K.** The sigature proteins of S2 subtype. **L.**The sigature proteins of S3 subtype.